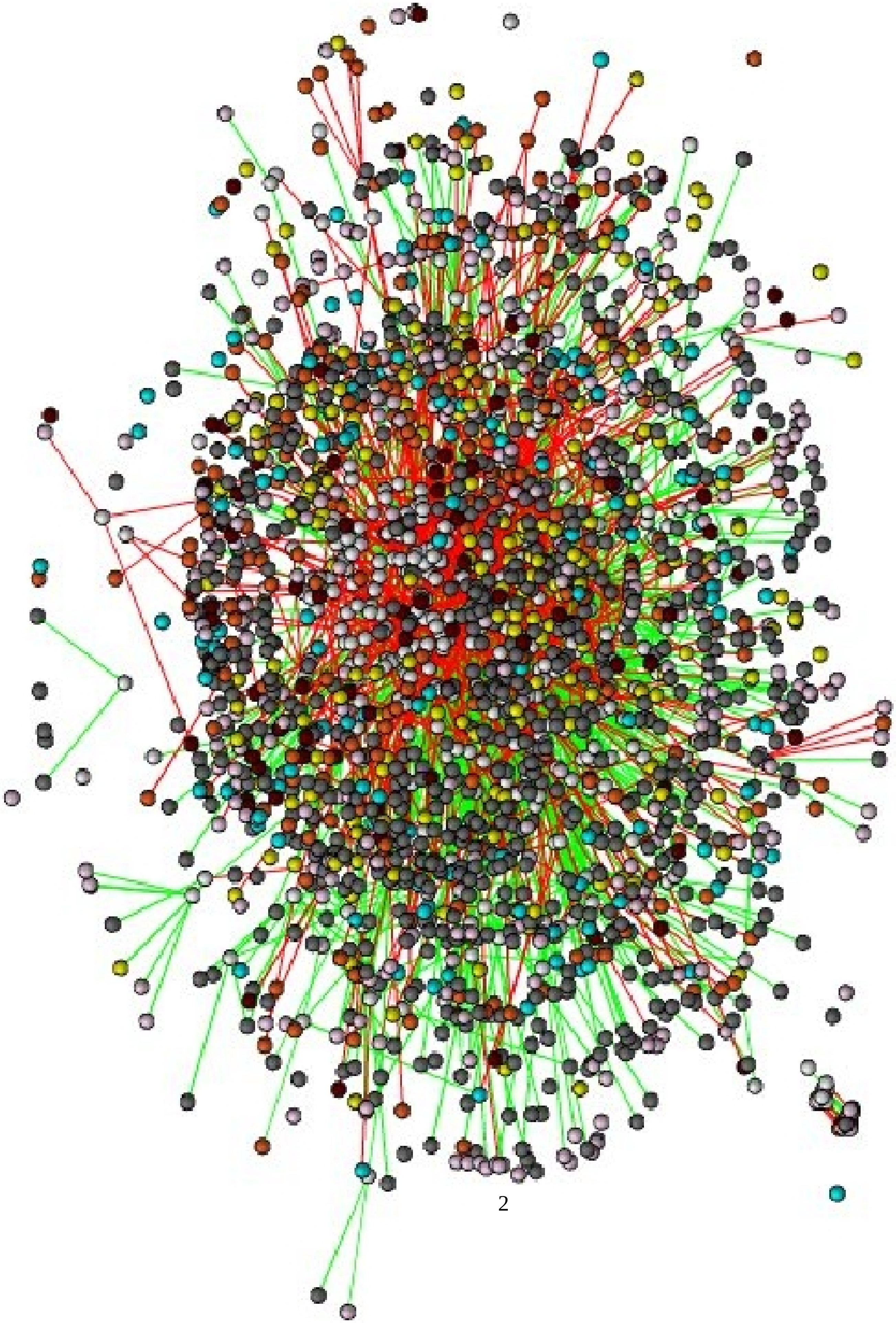
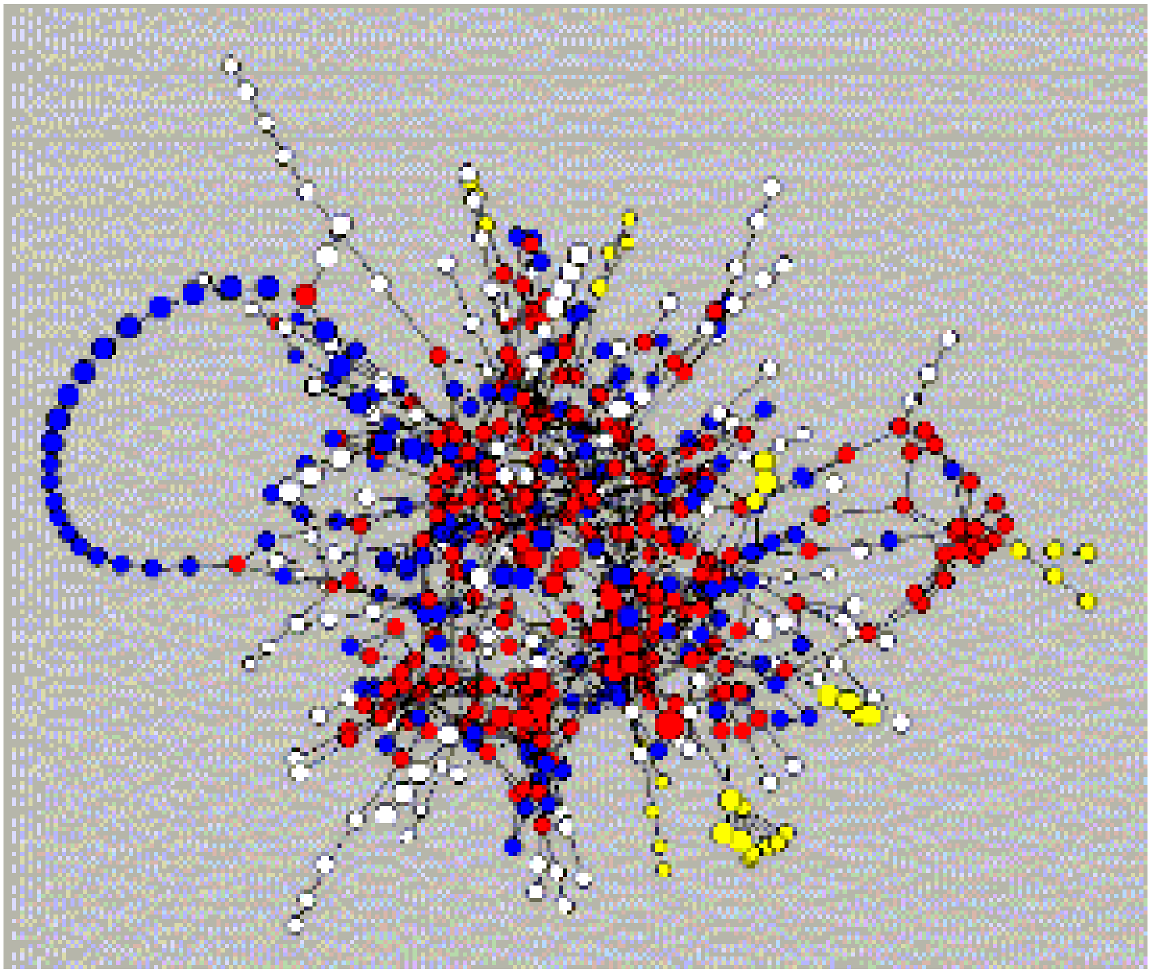
Scale-free networks: What are they and why should we care?

Rex K. Kincaid
Department of Mathematics
College of William and Mary

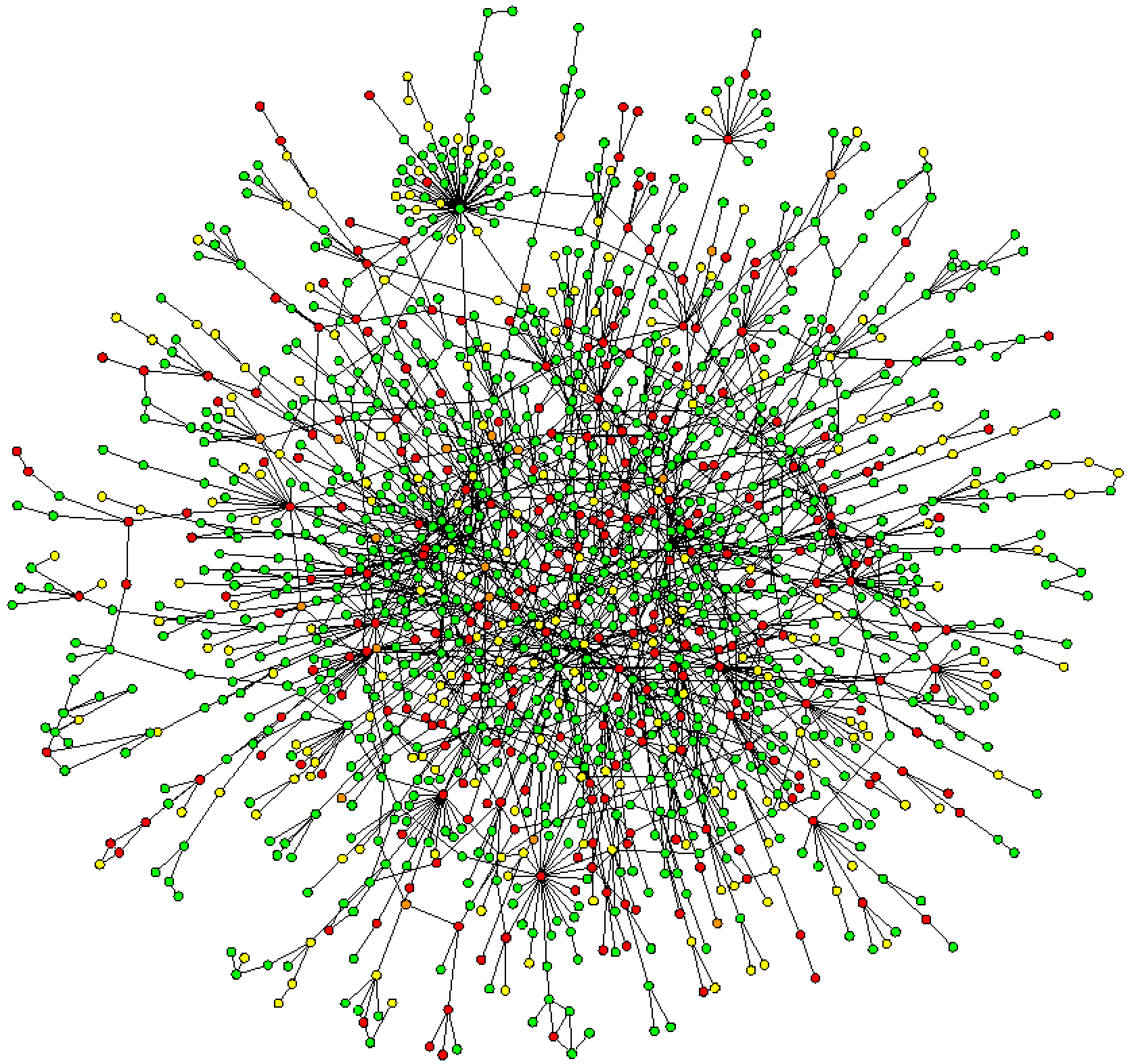
NASA Faculty Fellowship Program
July 25, 2003



Topology of the metabolic network of the yeast cell



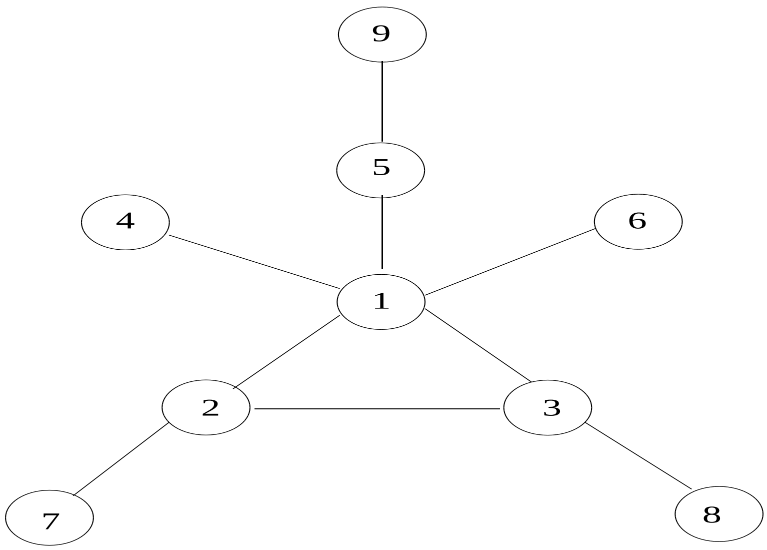
Map of protein-protein interactions



Random Graphs—$G_{n,p}$

- Erdos and Renya (1959,1960)
- n nodes
- p probability edge between two nodes
- $p \cdot n(n-1)/2$ expected number of edges
- degree distribution is binomial
- average shortest path $log\ n$
- clustering coefficient $(log\ n)/n$

Example: degree distribution



- degree of node 1 is 5
- degree sequence $\{1, 2, 3, 5\}$
- fraction of nodes of each degree k

  $p(k) = \{5/9, 1/9, 2/9, \ 1/9\}$
- fraction of nodes degree k or larger

  $P(k) = \{1, 4/9, 3/9, \ 1/9\}$—cdf

Clustering Coefficient

Transitivity: friend of your friend is also your friend
Goal: [0,1] number associated with number of triangles

Method 1:

$$C^{(1)} = \frac{\text{[three times total number of triangles]}}{\text{[total number of connected three tuples of nodes].}}$$

Method 2:

$$C_i = \frac{\text{[number of triangles in which node } i \text{ is incident]}}{\text{[number of three tuples of connected nodes centered on node } i\text{].}}$$

$$C^{(2)} = 1/n \cdot \Sigma_i C_i$$

- $C^{(1)} = 1/5$ and $C^{(2)} = 11/90$ in example

Properties of Scale-Free Graphs

- Small-world: average shortest path $log\ n$

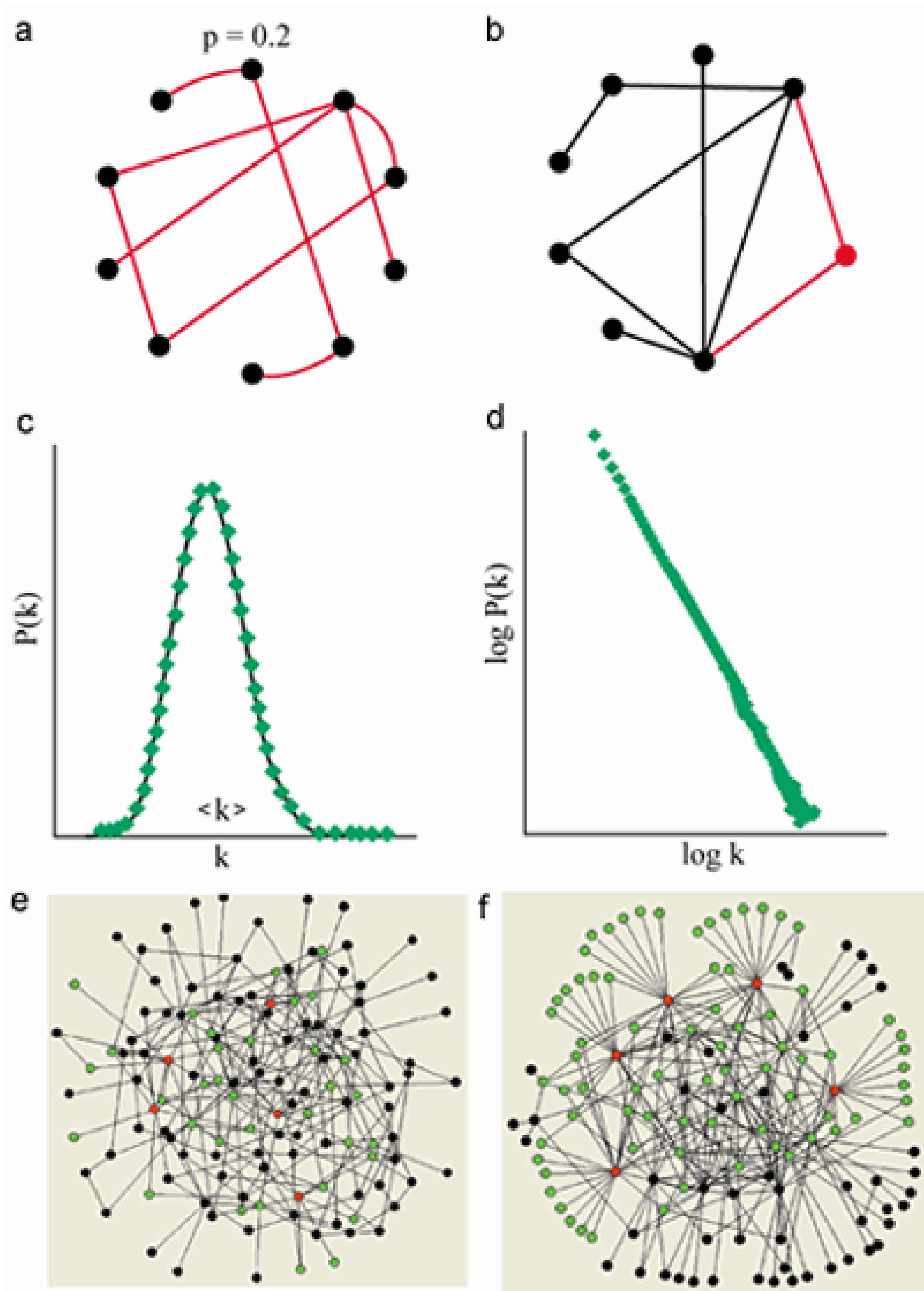
  Confusion in literature with graph diameter

- Power law: tail of $P(k)$ is linear log-log plot
- Exponential: $P(k)$ tail is linear for log plot
- Clustering Coefficient: constant for large n

Popularity of Scale-free Graphs

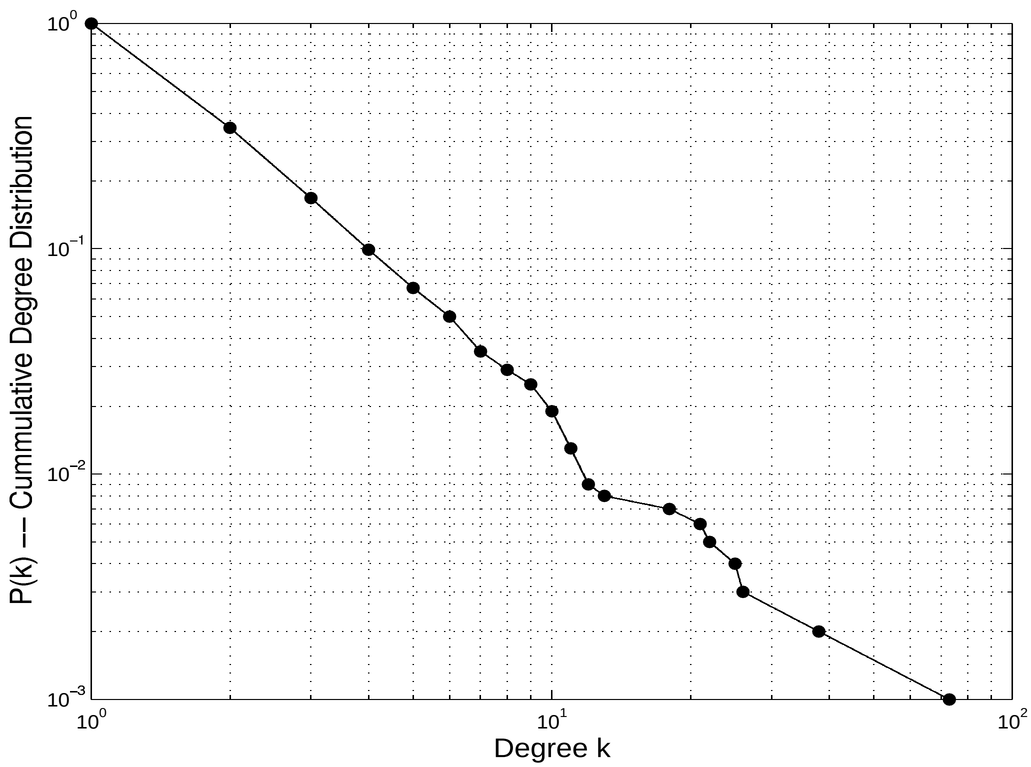
Graph	Nodes	Edges	AvgDeg	AvgSPD	$C^{(1)}$	$C^{(2)}$
film actor	449,913	25,516,482	113.443	3.48	.20	.78
Altavista	203,549,046	2,130,000,000	10.46	16.18	—	—-
Internet	10,697	31,992	5.98	3.31	.035	.39
protein	2,115	2,240	2.12	6.80	.072	.071

- 6 books—<u>Linked</u>, <u>Nexus</u>, <u>Six Degrees</u>.
- Review articles, articles in “Nature”, more than 500 research articles since 1997.
- Expected C value for random film actor graph is $.000252$.


a
p = 0.2
b
c
P(k)
<k>
k
d
log P(k)
log k
e
f

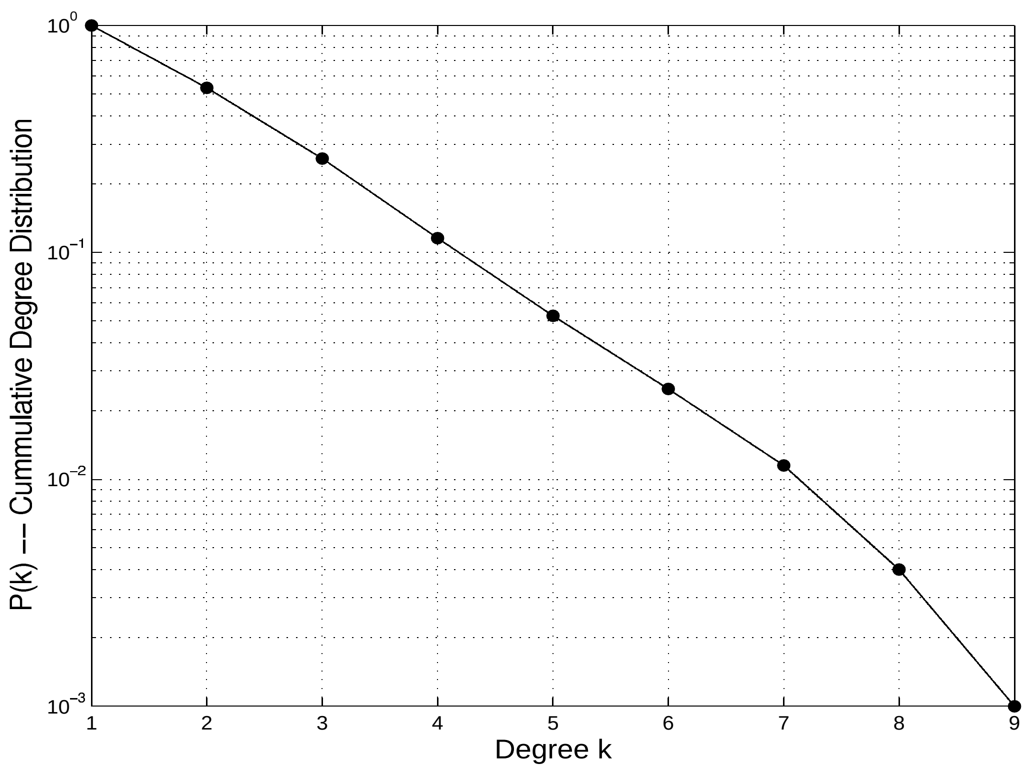
Scale-free graphs: preferential attachment

- Start with 3 node 2 edge path
- Add nodes one at a time: average degree
- Roulette wheel selection based on $p(k)$
- 1000 nodes



Scale-free graphs: Euclidean distance

- Start with 3 node 2 edge path
- Add nodes one at a time: average degree$= d$
- Pick d closest nodes (Euclidean distance)
- 2000 nodes; Western U.S. power grid



Strengths and Weaknesses

- **S**: small diameter (min communication costs)
- **W**: vulnerable to communicable diseases
- **W**: difficult to uncover shortest paths
- **S**: resistant to random node/edge failures
- **W**: powerless against targeted node attacks

How to strengthen scale-free graphs?

Callaway, D.S., M.E.J. Newman, S.H. Strogatz and D.J. Watts, "Network Robustness and Fragility: Percolation on Random Graphs," *Physical Review Letters*, **84** (2000) 5468-5471.

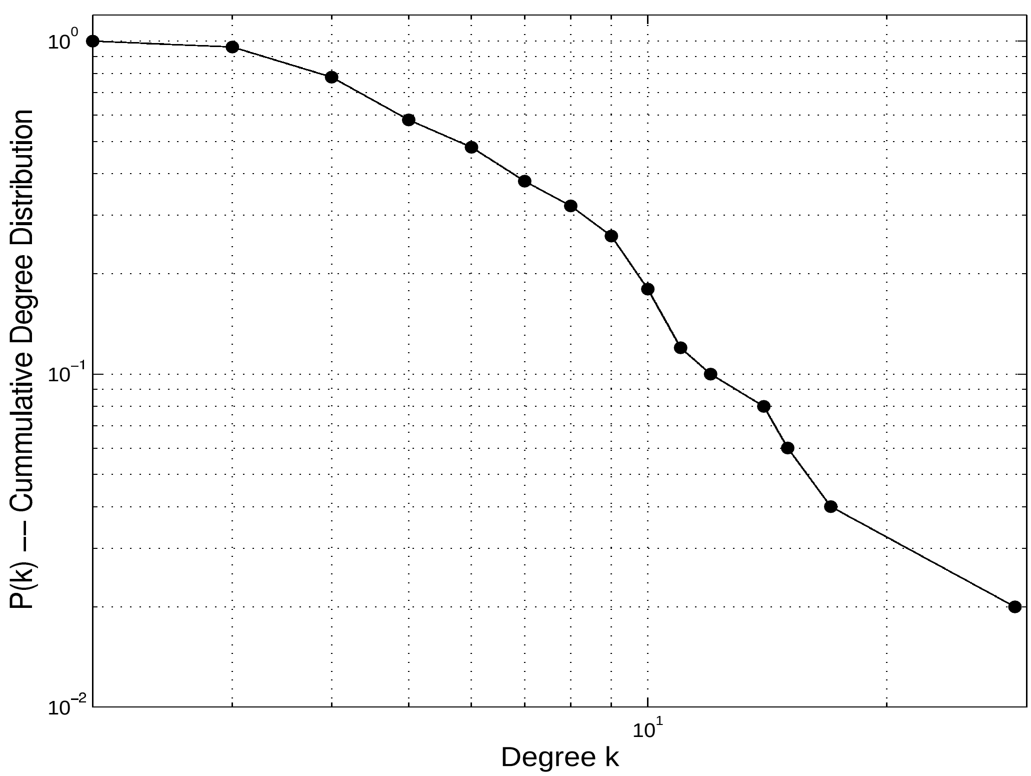
Park, S-T., A. Khrabrov, D.M. Pennock, S. Lawrence, C.L. Giles and L.H. Ungar, "Static and Dynamic Analysis of the Internet's Susceptibility to Faults and Attacks," *IEEE Infocom 2003*, Paper No. 0-7803-7753-2.

Optimization (tabu search) and scale-free

- Bi-objective models

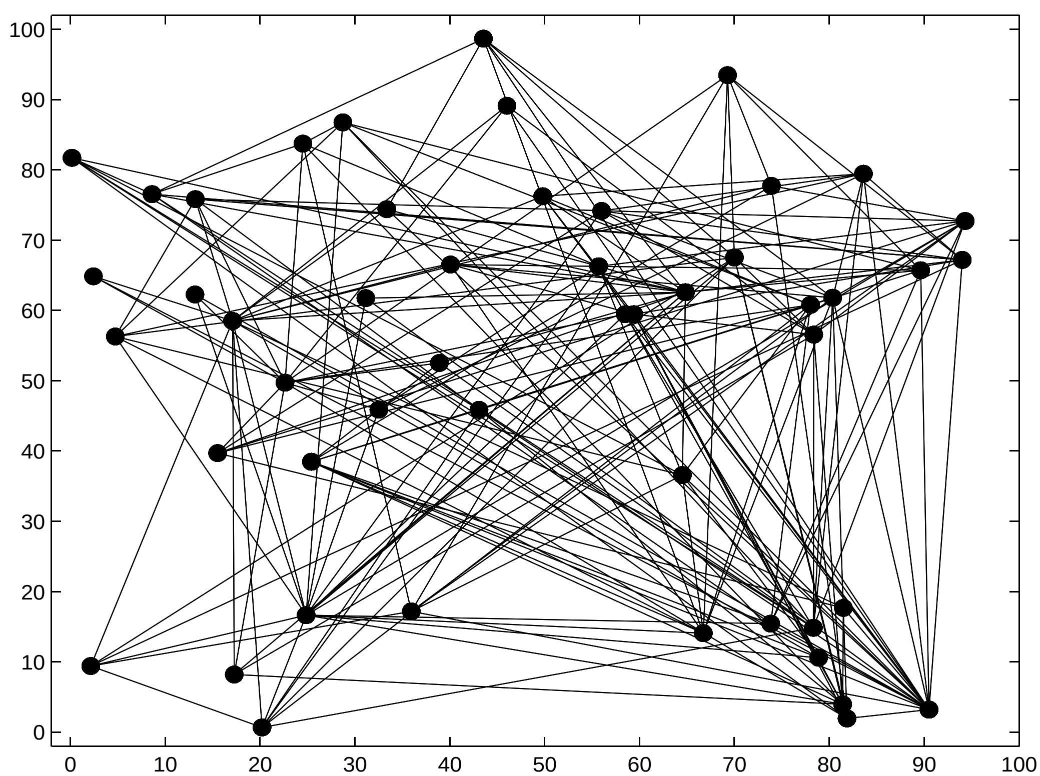
  Ferrer i Cancho and Solé (2001)—SFI.
  Min **(average shortest path + total number edges)**.
  Resulting graphs: trees to scale-free to star graph.

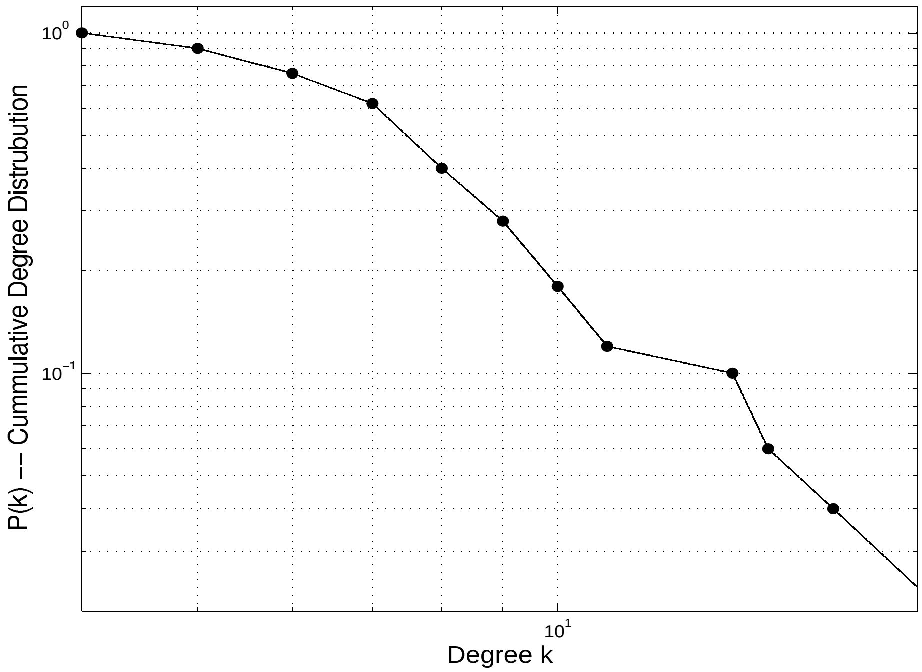
  (Plot below: 50 nodes, objectives equally weighted)



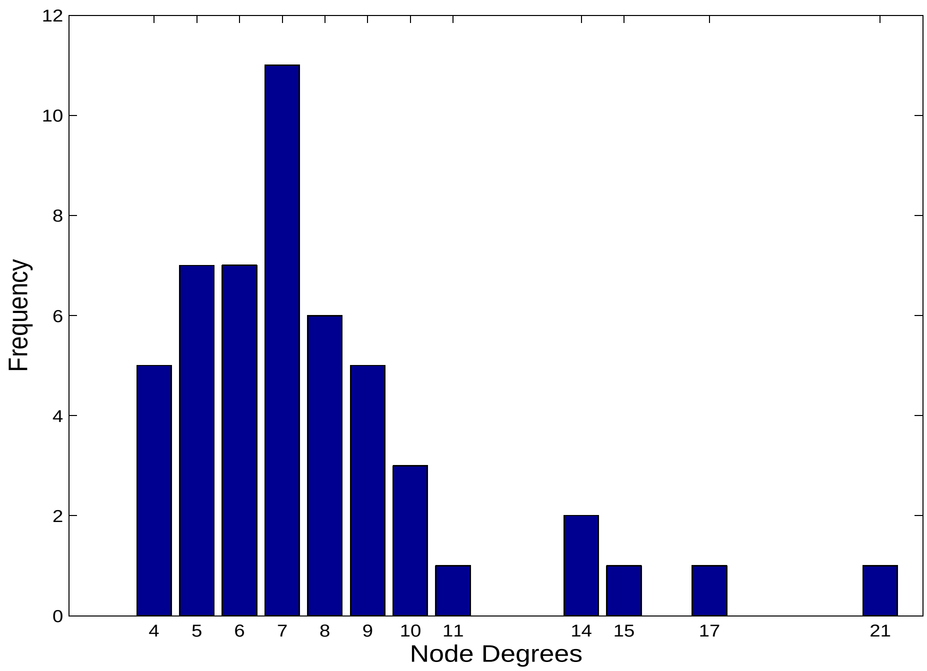
Optimization (tabu search) and scale-free

- Bi-objective models: replace **total number of edges** with **average degree**
- (Plot: 50 nodes, objectives weighted $.65/.35$)





10^0
10^−1
P(k) −− Cummulative Degree Distrubution
10^1
Degree k



12
10
8
6
4
2
0
Frequency
4 5 6 7 8 9 10 11 14 15 17 21
Node Degrees

How do scale-free graphs organize?

Ferrer and Sole (2001) believe that preferential attachment does not provide an adequate explanation for the large clustering coefficients observed in real networks. They believe that clustering is a side effect of optimization. They conjecture that reliable communication and cost minimizing shapes are the organizing principles behind scale-free graphs.

General Aviation Networks.

Lederer, P.J. and R.S. Nambimadom, “Airline Network Design,” Operations Research, **46** (1998) 785-804.

- Assumptions
  - — All cities served lie on a circle; equidistant
  - — Demand for service is identical for all O-D pairs
  - — Hub at circle center, not a destination
  - — Airline and passenger’s costs, 24 city max
- Network choice—hub and spoke, point to point, single tour, mixed subtours
- Sensitive parameters—demand, distance between cities, and number of cities
- Each network type optimal for some parameter choice.

General Aviation Networks.

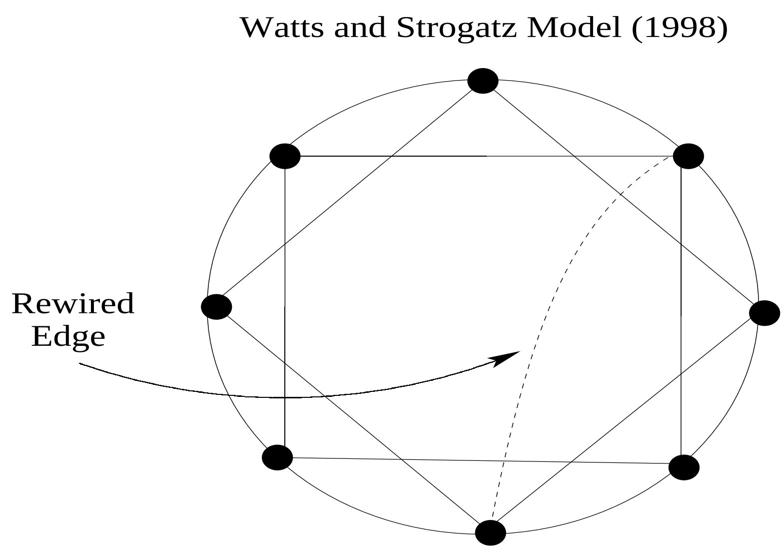
Yang, L. and R. Kornfeld, "Examination of the Hub-and-Spoke Network: A Case Example Using Overnight Package Delivery," in Proceedings of the 41st Aerospace Sciences Meeting and Exhibit, January 6-9, 2003.

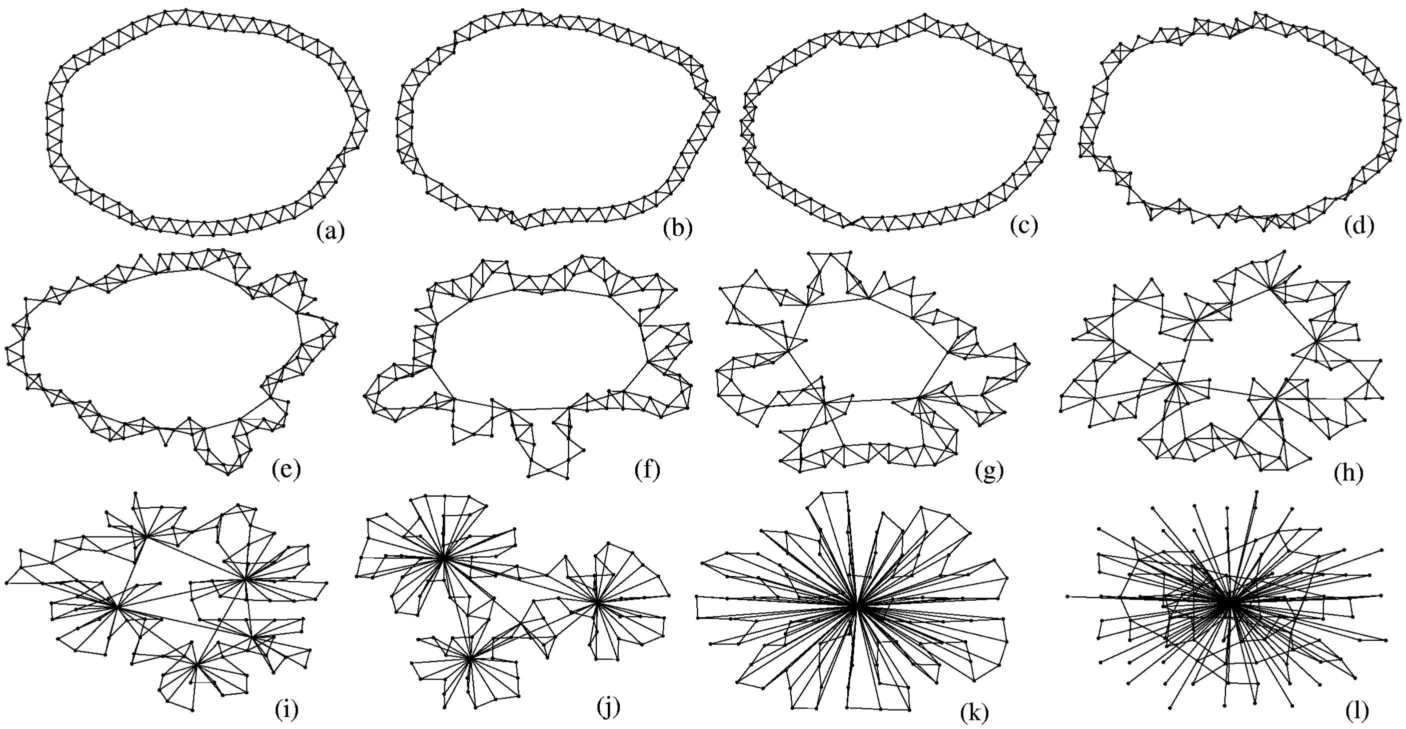
- Network choice—hub and spoke, point to point, several hubs
- Model—mixed linear integer program
- Limitations—only 7 city models tested
- Each network type optimal for some parameter choice
- LP-relaxation/column generation—more cities

Calgary ††
Vancouver
Seattle
Spokane
Kalispell
Havre
Glasgow
Wolf Point
Great Falls
Missoula
Lewistown
Sidney
Portland
Glendive
Helena
Eugene
Billings ††
Miles City
Medford †
Boise
Idaho Falls**
Salt Lake City
Reno
Sacramento
Oakland
San Francisco
San Jose
LAS VEGAS
Grand Junction
Aspen
Montrose
Telluride*
Durango
Denver
Colorado Springs
Farmington
Burbank
Ontario
Los Angeles
Long Beach
Orange County
San Diego
PHOENIX
Albuquerque
Tucson
El Paso
Hermosillo
Guaymas
Mazatlán
Los Cabos
Puerto Vallarta
Guadalajara
Manzanillo*
Ixtapa/Zihuatanejo
Acapulco
Mexico City
Minneapolis/St. Paul
Milwaukee
Des Moines
Omaha
Chicago (O'Hare)
Detroit
Cleveland
Toronto
Indianapolis
COLUMBUS
Pittsburgh ††
Kansas City
St. Louis
Wichita
Dallas/Ft. Worth (DFW)
Atlanta
Austin
San Antonio
Houston (IAH)
New Orleans
Orlando
Tampa
Fort Myers*
Fort Lauderdale/Hollywood
Miami
Raleigh/Durham
New York (LGA)
Newark
Hartford/Sp
Philadelphia
Baltimore (BWI)
Washington (Dulles)*
Washington (National)
Moloka'i
Maui
Lana'i
Kona
Hilo
The Big Island of Hawaii
Kingman
Flagstaff/ Grand Canyon
Lake Havasu City
Prescott
Yuma
PHOENIX
Sierra Vista/ Fort Huachuca
Stockton
Fresno
LAS VEGAS
Palm Springs
Carlsbad
PHOENIX
America West Express® service to Arizona and California cities.

Connections with Scale-Free Networks

- Start with k-regular ring lattice
- Rewire shortcuts across the lattice





(a) $\lambda=0$.(b) $\lambda=5\times10^{-4}$.(c) $\lambda=5\times10^{-3}$.(d) $\lambda=0.0125$.(e) $\lambda=0.025$.(f) $\lambda=0.05$.
(g) $\lambda=0.125$.(h) $\lambda=0.25$.(i) $\lambda=0.5$.(j) $\lambda=0.75$.(k) $\lambda=1$.

$\lambda\cdot$ (avg shortest path) $+(1-\lambda)$ (total distance)

General Aviation Problems

.

- Min cost solutions to daily routing problems
  - Assume routes are not fixed
  - Assume demand will increase
  - 3000 small/regional airports $<$ 2% of demand
- Test scale-free ring lattices against competing network types (Lederer 1998)
- Under what scenarios will scale-free win?
  - Include door-to-door costs?
  - What level fidelity needed to establish validity?
- Since regional airports are capcitated is scale-free limited?
- Seek local rules so that high quality route networks result and so that decision making is decentralized.

Related Combinatorial Optimization Problem

.

- Survivable network design problem
  - Undirected graph G with edge weights
  - Connectivity $r(i, j)$ specifiecd for all node pairs
  - Find minimum weight spanning subgraph with $r(i, j)$ edge-disjoint paths for all node pairs.
  - $r(i, j)$ provides measure of redundant connections to protect against edge failures.
- Problem is NP-hard (Steiner tree)
- Korte and Vygen (2000) reference
- Similar problem to address node failures? (dual)